

Opening a New Sport Unit in Berlin

Vera Rykalina
IBM by Coursera
January, 2020

Outline

Outline

Introduction

Data

Methodology

Results

Discussion

Recommendations

Conclusions

Introduction

Berlin City

Most valuable characteristics for living in Berlin is its ecological situation. Berlin is a green city. People in Berlin strive to lead a healthy life style. The inhabitants prefer clean eating and actively do sports.

In this report I would like to consider a possibility to open a new sport unit in Berlin preferably in the localities which are lacking sport centers. The bussiness decision should be made based on several parameters, mainly: a location, a population of the district and a number of the neighbouring sport units.

Bussiness Challenge

The objective of the given capstone project is to analyse and select the best locations in Berlin city to open a new sport unit (gym or fitness studio). Using data science methodology and machine learning algorithm such as clustering, the capstone project aims for providing solutions to answer the fowllowing business question: What are the best localities for a new sport unit in Berlin?

Target Audience

The given project can be particularly useful for a potential property developer or an investor who is planning to launch or invest in a new fitness unit in Berlin. The project goal is very challenging as the city is currently having a great number of sport spots but on the other hand, very fast growing dynamic and active population of Berlin is demanding in terms of variety of sport activities and vacant memberships in already existing sport centers. Besides, there are localities, which have a limited number of the sport schools and centers due to historical or economical reasons but still with increasing demand for such units.

Data

Type of Data

- A list of the localities in Berlin.
- Demographic characteristics of each locality.
- Latitude and longitude coordinates of the localities. This type of data is required for locality mapping and venue data establishment.
- A venue data preferentially anchored in the data related to a gym or a fitness center for clustering.

Data Mining

- Wikipedia page for Berlin localities and neighbourhoods
- Geocoder package for geographical coordinates
- Foursquare API for venue data

Methodology

Techniques

- Web scrapping Wki page for gaining a list of the localities
- Latitude and Longitude coordinates acquisition by means of Geocoder
- Utilizing Foursquare API for obtaining venue data
- Grouping data by locality and taking the mean of the category occurrence frequency
- Filtering venue category against fitness studio
- Clustering performance using K-means algorithm
- Mapping clusters using Folium

Results

Clusters

- Cluster 2: Localities with moderate number of fitness centers
- Cluster 1: Localities with low number to no existence of fitness centers
- Cluster 0: Localities with high concentration of fitness centers

Cluster Mapping

The result from the k-means clustering demonstrates clear distribution of clusters based on different frequency of occurrence of the gym/fitness studios

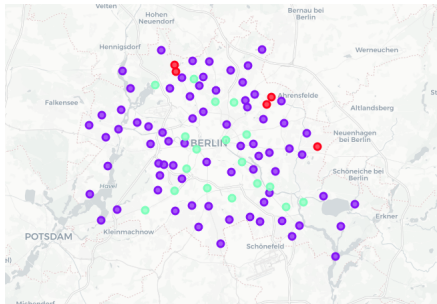


Figure 1: Mapping of Clusters for localities in Berlin

Discussion

Key Findings

Most of the gyms or fitness centers with moderate concentration are located in the city center. The highly concentrated sport spots (cluster 0) takes place in the north-east part of Berlin. These sport units were formed due to geographical convenience and historical reasons. On the other hand, cluster 1 has a very low number to no fitness centers in the locality what can be interpreted as a great opportunity for launching a new sport unit. Besides, sport unit in these localities are far away from the city center and could be a reasonable investment.

Recommendations

Capstone project recommendations to property developers to make their investments based on the clustering findings:

- Investors are advised to consider launching a new sport unit in any locality in cluster 1 with low competition level and lower building price per square meter than in the city center.
- Stakeholders are also encouraged to look for opportunities to open new sport unit in any locality in cluster 0, if they are ready for a moderate competition conditions.
- Potential property developers should avoid localities in cluster 2 which already have high concentration of gyms and fitness studios.

Conclusions

Conclusions

- In the given project several steps were taken for stating a business problem, specifying the required data, extracting and preparing the data and performing machine learning clustering algorithm to group the localities into 3 clusters based on their similarities.
- Moreover, the useful recommendations were provided to the potential stakeholders i.e. property developers and investors regarding the best locations to open a new fitness center. Back to the business question framed in the introduction section, the localities in cluster 1 are the most preferred locations to open a new gym or a fitness center.
- The findings of this project can be helpful to relevant stakeholders to make their investments in the reasonable way and get payoffs in short period of time.

Thank you!